

# Machine Learning 1

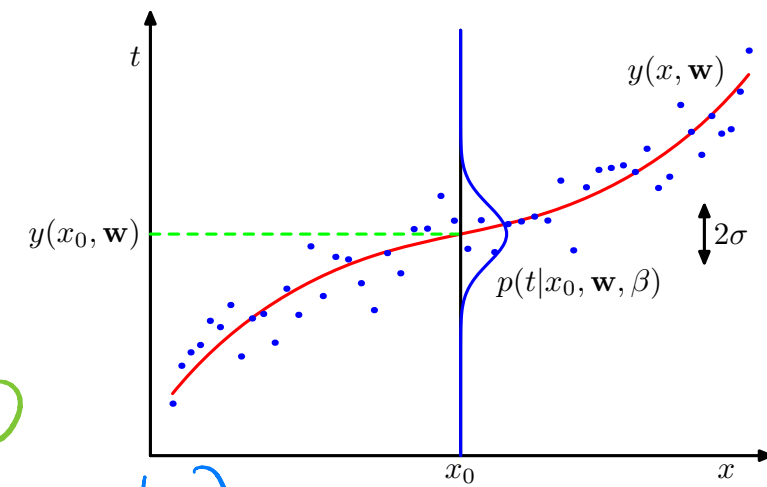
Lecture 4.3 - Supervised Learning  
Bayesian Linear Regression - Gaussian  
Posteriors

*Erik Bekkers*

*(Bishop 3.3.1 (and 2.3.3))*



# Bayesian Linear Regression



► Data:  $\mathbf{t} = (t_1, \dots, t_N)^T$   $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)^T$

► Likelihood:  $p(t' | \mathbf{x}', \mathbf{w}, \beta) = \mathcal{N}(t' | \underbrace{\mathbf{w}^T \phi(\mathbf{x}')}_{y(\mathbf{x}', \mathbf{w})}, \beta^{-1})$

very

$$p(\mathbf{t} | \mathbf{X}, \mathbf{w}, \beta) = \prod_{n=1}^N \mathcal{N}(t_n | y(\mathbf{x}_n, \mathbf{w}), \beta^{-1}) = \mathcal{N}(\underline{t} | \Phi \underline{w}, \beta^{-1} \mathbf{I})$$

► Conjugate prior:  $p(\mathbf{w}) = \mathcal{N}(\mathbf{w} | \mathbf{m}_0, \mathbf{S}_0)$   $\Phi = \begin{pmatrix} \phi_0(\mathbf{x}_1) & \dots & \phi_{M-1}(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ \phi_0(\mathbf{x}_N) & \dots & \phi_{M-1}(\mathbf{x}_N) \end{pmatrix}$

► Posterior distribution:

$$p(\mathbf{w} | \mathbf{t}, \mathbf{X}) = \frac{p(\mathbf{t} | \mathbf{X}, \mathbf{w}, \beta) p(\mathbf{w})}{p(\mathbf{t} | \mathbf{X}, \beta)} = \mathcal{N}(\underline{w} | \underline{m}_N, \mathbf{S}_N)$$

Ch 2.3

Bishop Eq. 2.116

► Maximum A Posteriori estimate:

$$\mathbf{w}_{\text{MAP}} = \underline{m}_N$$

$$\begin{aligned} \mathbf{S}_N^{-1} &= \mathbf{S}_0^{-1} + \beta \Phi^T \Phi \\ \mathbf{m}_N &= \mathbf{S}_N (\mathbf{S}_0^{-1} \mathbf{m}_0 + \beta \Phi^T \mathbf{t}) \end{aligned}$$

# Bayesian Linear Regression

- Special simple prior:  $p(\mathbf{w}|\alpha) = \mathcal{N}(\underline{\mathbf{w}}|\underline{\mathbf{0}}, \alpha^{-1}\mathbf{I})$   $\mathbf{m}_0 = \mathbf{0}$   
 $\mathbf{S}_0 = \alpha^{-1}\mathbf{I}$

- Posterior  $p(\mathbf{w}|\mathbf{X}, \mathbf{t}, \alpha, \beta) = \mathcal{N}(\mathbf{w}|\mathbf{m}_N, \mathbf{S}_N)$

$$\mathbf{m}_N = \mathbf{S}_N(\cancel{\mathbf{S}_0^{-1}\mathbf{m}_0} + \beta\Phi^T\mathbf{t}) = \beta\mathbf{S}_N\Phi^T\mathbf{t}$$

$$\mathbf{S}_N^{-1} = \mathbf{S}_0^{-1} + \beta\Phi^T\Phi = \alpha\mathbf{I} + \beta\Phi^T\Phi$$

$$p(\mathbf{w}|\mathbf{X}, \mathbf{t}, \alpha, \beta) = \frac{1}{\sqrt{(2\pi)^M \cdot |\mathbf{S}_N|}} e^{-\frac{1}{2}(\underline{\mathbf{w}} - \underline{\mathbf{m}}_N)^T \mathbf{S}_N^{-1} (\underline{\mathbf{w}} - \underline{\mathbf{m}}_N)}$$

# Bayesian Linear Regression

**Limiting cases:**  $p(\mathbf{w}|\mathbf{X}, \mathbf{t}, \alpha, \beta) = \mathcal{N}(\mathbf{w}|\mathbf{m}_N, \mathbf{S}_N)$   $\mathbf{m}_N = \beta \mathbf{S}_N \Phi^T \mathbf{t}$   
 $\mathbf{S}_N^{-1} = \alpha \mathbf{1} + \beta \Phi^T \Phi$

- ▶ Infinitely broad prior: no restriction on  $\mathbf{w}$ !

$$p(\mathbf{w}|\alpha) = \mathcal{N}(\mathbf{w}|0, \alpha^{-1} \mathbf{1}) \quad \alpha \rightarrow 0$$

$$\lim_{\alpha \rightarrow 0} \mathbf{m}_N = \lim_{\alpha \rightarrow 0} \beta \left( \alpha \mathbf{1} + \beta \Phi^T \Phi \right)^{-1} \Phi^T \mathbf{t} = \frac{\beta}{\beta} (\Phi^T \Phi)^{-1} \Phi^T \mathbf{t} = \mathbf{w}_{ML}$$

- ▶ Infinitely narrow prior:

$$p(\mathbf{w}|\alpha) = \mathcal{N}(\mathbf{w}|0, \alpha^{-1} \mathbf{1}) \quad \alpha \rightarrow \infty$$

$$\lim_{\alpha \rightarrow \infty} \mathbf{m}_N = \lim_{\alpha \rightarrow \infty} \beta \left( \alpha \mathbf{1} + \beta \Phi^T \Phi \right)^{-1} \Phi^T \mathbf{t} = \lim_{\alpha \rightarrow \infty} \frac{\beta}{\alpha} \Phi^T \mathbf{t} = \mathbf{0} = \mathbf{m}_N$$

$$\lim_{\alpha \rightarrow \infty} \mathbf{S}_N = \lim_{\alpha \rightarrow \infty} \left( \alpha \mathbf{1} + \beta \Phi^T \Phi \right)^{-1} = \lim_{\alpha \rightarrow \infty} \frac{1}{\alpha} \mathbf{I} = \mathbf{0}$$

zero matrix